PRELIMINARY DRAFT

The Effects of Neighborhood Quality During Childhood On Health & Health Behaviors in Early and Mid Life

Rucker C. Johnson*
Robert Wood Johnson Postdoctoral Research Fellow
University of Michigan

Robert F. Schoeni
Senior Associate Research Scientist, Institute for Social Research
Associate Professor of Economics & Public Policy
University of Michigan

First Draft: April 2003

Paper prepared for presentation at the RAND Health Economics Conference, April 27-28, 2003, Santa Monica, CA.

* Please direct correspondence to Rucker Johnson, University of Michigan, 109 Observatory, SPH-II, M2226, Ann Arbor, MI 48197, or email to ruckerj@umich.edu. This research is supported by funding from the Robert Wood Johnson Foundation to Johnson and NIA grant K01AG00670 to Schoeni.

I. INTRODUCTION

The persistent residential segregation of poor and minority populations has spurred a growing body of literature on inequality that investigates the effects of community background on a variety of socioeconomic outcomes. However, the effects of the physical and socioeconomic neighborhood environment on health outcomes has been relatively unexplored.

The health inequality literature has focused largely on individual-level determinants and family background effects on health behaviors and health outcomes. Studies have documented the importance of such factors as age, marital status, education, income, wealth, and health care utilization patterns. However, these factors collectively explain only a small proportion of the total variation in health outcomes and health-related behaviors.

At the same time, health outcomes and behaviors exhibit a distinctive spatial pattern that mirrors the spatial pattern of physical and socioeconomic disadvantage. The similarity of these geographic patterns motivates our investigation into the potential causal role of neighborhood context on health and health behaviors. Specifically, in this paper we analyze the effects that the physical and socioeconomic neighborhood environment in which an individual grows up has on their subsequent health and health-related behaviors. Using data from the Panel Study of Income Dynamics spanning 1968 to 2001, we analyze five dimensions of health over the life-course:

1. Low birth weight;
2. Poor health status in childhood and in early to mid adulthood;
3. Obesity in early to mid adulthood;

4. Age of onset of cigarette smoking, and smoking behavior in adulthood;
5. Risk preferences.

Our analysis of correlations between siblings and between neighboring children in their subsequent health outcomes in adulthood provides new insights about the potential scope of neighborhood effects on health over the life course. Our results also suggest neighborhood factors play an important role in the intergenerational transmission of health status. The evidence presented in this paper provides direction for future research to further our understanding of the underlying processes that produce health disparities between different racial/ethnic and socioeconomic groups.

The paper is organized in the following way. We begin with a discussion of why neighborhood context may matter for health-related outcomes. We outline an economic model of health that incorporates the influence of neighborhood factors. The model provides our theoretical framework, highlights the relevant theoretical issues, and motivates the empirical analyses to follow. Section III lays out the methodological challenges in estimating neighborhood effects. The data and outcome variables are described in section IV. Sections V and VI describe the econometric model and estimation methods, respectively. The results are presented in section VI, with concluding statements provided in the final section.

II. WHY MIGHT NEIGHBORHOODS MATTER?

The physical and socioeconomic context of communities may influence the health of individual residents over and above their own socioeconomic position in many plausible ways.[1] We will describe an economic model of health that incorporates the

[1] Recent literature reviews of neighborhood effects on health include Ellen, Mijanovich, Dillman (2001); Diex-Roux (2001); Morenoff & Lynch (2002); Pickett & Pearl (2001); Robert (1999); Yen & Syme (1999).

potentially causal role of neighborhood context, including causes related to 1) the composition and behavioral choices of its members such as peer group effects, role model effects, and effects of social complementarities; 2) the differences in contextual-level neighborhood conditions such as housing quality, over-crowding, deteriorating infrastructures, sanitation, water and air quality, exposure to environmental toxins, pollutants, and industrial hazards; and 3) the differences in institutional quantity and quality of community resources, most importantly access to quality and quantity of health care and social services (Chandra & Skinner 2003, Skinner et al. 2002).

Our formal model builds from standard microeconomic foundations and integrates sociological and psychological perspectives. As in standard economic models of behavior, individuals make purposeful decisions based on their preferences, their beliefs about the consequences of alternative actions, and constraints that delimit those actions. The main theoretical extension and modeling insight that motivate empirical analyses of neighborhood effects is to explore how neighborhood context influences preferences, beliefs, and constraints. We refer to this as endogenous preference formation—i.e., how space and community influence individual perceptions, aspirations, and opportunities. This represents a significant departure from simplistic assumptions embodied in Tiebout sorting of exogenous preferences into neighborhoods with homogeneous preferences within neighborhoods.

The model has the capacity to explain the emergence of clustering of a wide array of seemingly different phenomena, such as poor health outcomes, health-related behaviors, substance abuse, educational attainment, age of onset of adolescent sexual activity, sexually-transmitted disease, out-of-wedlock births, criminal activity, low

earnings, low wealth holdings and savings behavior, and entrepreneurial activity with a common underlying process. The key insight is that all of these choices can be viewed in an investment framework—whether health investment, human capital investment, or financial investment. Thus, it is important to investigate the effects of neighborhood context in shaping risk-taking, rate of time preference, and intertemporal substitution parameters because these parameters are determinants of these outcomes. Simple "culture of poverty" explanations in which individuals do not respond to incentives and opportunities are incomplete because the ways in which neighborhood context shapes preferences, beliefs, and constraints are not considered.

THEORETICAL MODEL/FRAMEWORK

Building on the models of Dow et al. (1999), Ganz (2001), and Durlauf (2000), consider a utility-based framework, where the health behavior choices of each individual i in neighborhood n represent the solution to:

$$\max_{B_i \in \Omega_i} \; V(H_i,(B_i - m_{n(i)}^e),C_i)$$

such that $H_i=H(B_i,E_{n(i)},X_i)$

such that $\Omega_i=\Omega(X_i, E_{n(i)},\mu_i^e(B_{n(-i)}),C_i)$

where V is the utility function; H is the health production function; B are health behaviors; $m_{n(i)}^e$ is individual i's expectation of the mean health behavior of others in the neighborhood; C is consumption of all other goods; E is environment/neighborhood conditions; X are heterogeneous individual traits; $\mu_i^e(B_{n(-i)})$ represent individual i's expectation of the health behaviors of others in the neighborhood; and choices lie within the constraint set Ω_i each individual faces. Assume investments in health and improvements in neighborhood conditions have positive effects on the production of

health ($H_B>0$, $H_E>0$), and assume this productivity exhibits decreasing returns ($H_{BB}<0$, $H_{EE}<0$).

Of particular interest are the interactions between neighborhood influences and behaviors. First, we model utility as a function of $(B_i - m_{n(i)}^e)$ to capture a type of peer group effect, reflecting the disutility from failing to conform to the expected average behavior of others in the neighborhood ($H_{(B_i - m_{n(i)}^e)} < 0$). In essence, this is the impact the choices of some neighbors have on the preferences of others in assessing those same choices. For example, in the case of cigarette smoking, use of cigarettes is more appealing when one's friends also smoke, particularly in adolescence, when the onset of these addictive behaviors occurs. We will present empirical results in Section VII that attempt to shed light on the potential strength of this type of peer effect.

Second, we hypothesize that health behaviors and the quality of neighborhood conditions are complementary inputs in the production of health, which would imply the cross-partial derivative $H_{BE}>0$. These complementarities follow from the competing risk model formalized by Dow et al. (1999) where overall life span (or healthy life years), H, is a function of competing risks which takes the form, $H=f(t_1, t_2,\ldots,t_k)=min(t_1, t_2,\ldots,t_k)$, where t_i is the life span associated with mortality (or morbidity) risk r_i. In other words, the most immediate threat to life and health determines an individual's overall life span (or span of healthy time). This directly implies that the competing risk model is a Leontief-type health production function, which further implies positive complementarities among the inputs, taken here to be preventive measures against the k health threats.

These risks include poor neighborhood conditions such as exposure to environmental toxins, pollution, poor air or water quality, poor housing quality, inferior access to quantity and quality of health services, high neighborhood rates of crime and violence, which all may affect actual life expectancy as well as perceived risk and perceived life expectancy. As Dow et al. (1999) show, the marginal benefit of resources directed at a health threat is zero if that threat is not the most immediate.[2] Health production complements imply demand complements (Ganz 2001), highlighting the potential causal role of neighborhood conditions on health behavior. That is, neighborhood conditions have an indirect effect on health through their effects on health behaviors, as well as its more commonly cited direct effect on health through "weathering," whereby the accumulated stress, lower environmental quality, and limited resources of poorer communities experienced over many years erodes the health of residents in ways that make them more vulnerable to mortality from any given disease (Geronimus 1992). Individuals evaluate their own lifetime survival probabilities, and external health threats lower these survival probabilities. Individuals who face a shortened life span because of poor neighborhood conditions have weaker incentives to engage in preventive health behaviors that benefit life span and health in the long-run. The model, thus, predicts improvements to the physical and social environment will lead individuals to invest more in private health production because they have a longer perceived life span over which to reap the benefits of healthier living.

[2] The deterministic case generates preferences where a person only cares about the risks in order of their intensity. Relaxing this assumption by assuming that each risk is a hazard rate would generate some smoothness in individual preferences (Dow et al. (1999); Ganz (2001)). Individuals would care more about the risks that have larger hazard rates, but would still care some about all risks since each has a positive probability to bind first. Dow et al. (1999) show that the same behavioral implications hold in both the deterministic and stochastic cases.

Smoking behavior, which we examine empirically, provides a good illustrative example of potential effects of neighborhood context on health behavior. Cigarette smoking has immediate benefits as a coping strategy for the temporary relief of life stressors, some of which may result from neighborhood deprivation. But smoking is accompanied by long-run costs in the form of increased health hazards. In the competing risk model, external health threats such as exposure to high levels of neighborhood violence and crime will change the perceived risks of such unhealthy behaviors as smoking because the presence of more immediate health threats make longer-term health dangers seem more remote. Thus, adoption of health-threatening behaviors may be a rational choice in neighborhoods where expected life-span is relatively short because the health costs of such behaviors is overshadowed by other, more immediate threats to health, and residents have less to lose from health risks such as smoking, that typically take a long time to manifest.

This model also underscores the potential causal role of childhood neighborhood conditions on individual risk preference formation, the rate of time preference, and intertemporal substitution parameters, more generally. In an investment framework, smoking behavior can be viewed as just one of the many risky behaviors that neighborhood context influences through its affect on risk preference formation. Empirically we will investigate this by analyzing correlations in risk tolerance in adulthood between siblings and between childhood neighbors.

The model demonstrates that the neighborhood in which children grow up can have a strong effect on future health and health behaviors. But a complete neighborhood effects theory also requires an explanation of how neighborhoods form. Neighborhoods

can be viewed as a byproduct of, among other things, racial and economic segregation which emerges because families prefer affluent neighbors due to their effects on the tax base, and families prefer affluent neighbors due to the role model influences that they produce. Schelling (1971) has shown in theoretical models that even "mild" preferences to be in an ethnic majority in their communities can produce very segregated communities. Furthermore, Durlauf and others have shown that sorting based on these processes is not necessarily socially efficient. Low-income households and minorities face residential location constraints due to suburban land use practices such as exclusionary zoning and discrimination (Yinger, 1995). The existence of these residential location constraints undermines the simple story that observed geographic clustering of people with similar health outcomes is the result of Tiebout sorting of exogenous preferences.

Peer group effects, role model effects, and contextual-complementarity effects, described above, each represent distinct influences under the umbrella of neighborhood effects. In the analyses conducted in this paper, we will not attempt to decompose these separate sources of neighborhood effects, but rather first focus on quantifying the potential overall magnitude of neighborhood effects. This decomposition and investigation into the mechanisms of why neighborhoods matter are an important next step and area for future research.

III. Methodological challenges in Estimating Neighborhood Effects

The primary methodological challenge in estimating the causal effects of neighborhoods on health and health-related behaviors is that unobserved factors that affect health may also be correlated with neighborhood factors, leading to biased

estimates of neighborhood effects. This can arise due to the endogeneity of residential location. That is, individuals and families choose where they live based on the characteristics they value (Tiebout 1956)—although, residential segregation by race and income and suburban land use policies, such as exclusionary zoning, may constrain the residential location choices of low-income households and minorities. In this context, families and individuals who care more about their health will be less likely to choose to live in an area with high crime, pollution, or a poor health care system. As a result, the systematic sorting of more health-conscious individuals (who consequently have healthier behaviors) into neighborhoods that are supportive of good health causes estimates of neighborhood effects to be biased upwards since the selection effects along these unobserved dimensions cannot be fully controlled. Furthermore, estimates of neighborhood effects may suffer from omitted variable bias because the neighborhood variables are serving as proxies for unmeasured aspects of family background that cannot be fully controlled.

The most powerful way to address selection is through a randomized trial. But an experimental design where neighborhoods are randomly assigned is very rare, and the few studies that have been conducted (the Gautreaux program in Chicago and the Moving to Opportunity program) are not well designed to examine the long-run effects of neighborhood conditions on health, or differential effects over the life-course, nor are they immune from problems with selection.

Among the studies that have tried to address endogeneity and self-selection, the most common approach is the use of instrumental variable techniques (e.g., Evans et al 1992; Case and Katz 1991; and McLanahan 1996). The goal of this approach is to

identify at least one variable that affects the choice of neighborhood but does not affect the outcome of interest, in our case health status in early to mid adulthood. However, finding a valid instrument is difficult if not impossible and often relies on strong priors about the correlation between the instruments and the error term in the health status equation.

An alternative non-experimental approach is comparing siblings who have been raised in different neighborhoods at different ages because their parents have moved (Aaronson 1997, 1998; Plotnick and Hoffman 1996). The rationale for this approach is that the neighborhood choice is determined by a fixed family factor and then a variety of other factors that are out of the control of the family, like changes in the labor market or health shocks of extended family members. The key assumption is that the family effect is fixed – not time-varying. If, for example, families preferences change as their children get older, and they become more interested in living in neighborhoods that are less risky for their children's health, then they might move to neighborhoods with less crime or pollution, which my in turn lead to better health outcomes for their kids. But if the underlying change in their preferences towards enhanced health outcomes not only caused them to change neighborhoods, but also to spend more time encouraging their children to practice good health behaviors such as eating healthily, exercising, and avoiding high crime areas, then the neighborhood "effect" might actually be representing all of these other factors and not the true causal effects of neighborhoods *per se*. Moreover, it is quite possible that sibling differences may aggravate the endogeneity problem, as has been discussed in the context of the labor market returns to schooling (Griliches 1979; Bound and Solon 1999).

An additional estimation issue that arises is related to the difficulty in appropriately measuring neighborhood conditions. The neighborhood factors that may in fact matter may be hard to measure, or they may not be measured in enough spatial detail. This issue is analogous to the finding in the family background literature that sibling correlations in socioeconomic status far exceed what has been explained by any particular measured aspects of the siblings' shared background (Corcoran, Jencks, and Olneck 1976).

Instead of performing another regression analysis focused on particular neighborhood characteristics, in this paper we exploit a unique feature of the Panel Study of Income Dynamics (PSID) and adopt an approach recently used by Solon et al (2000) to examine the effects of neighborhood factors on educational attainment. Specifically, the initial PSID sample in 1968 was highly clustered with most PSID families having several other sample families living in the same block. This design allows us to compare the similarity in early to mid adulthood health between siblings who grew up together, versus unrelated individuals who grew up in the same narrowly defined neighborhood. We use correlations between neighboring children's subsequent health in early to mid adulthood to bound the proportion of inequality in health outcomes that can be attributed to disparities in neighborhood background. Because the PSID sampled neighboring children and all children within 1968 family members, and then followed them into adulthood, we use sibling correlations in early adulthood health outcomes and correlations between unrelated neighbors to assess the relative correlation of neighborhood versus family background factors. This approach avoids the difficulty of

defining neighborhood quality and instead compares sibling correlations with neighbor correlations.

IV. Data and Health Outcomes

Our data on siblings and childhood neighbors come from the PSID, a nationally representative longitudinal survey conducted by the University of Michigan's Institute for Social Research. The PSID began by interviewing a national probability sample of families in 1968 and has re-interviewed the members of those families every year since. The PSID used a "cluster sample" (i.e., several households were selected in the same vicinity, usually within a block or two of each other) when it started in 1968 in order to economize on interviewing costs. This design effect is typically a liability in statistical analyses because one has to account for non-independence across individuals within the same cluster. But for our purposes the clustering provides the unique opportunity to examine health outcomes for adults who were childhood neighbors in 1968. Moreover, because all 1968 family members within a given family are followed throughout their lives, we can examine the similarity in health outcomes over the life-course of both siblings and childhood neighbors.

In our analyses, we define the neighborhood of upbringing as the census tract where the child lived in 1968. Not all parts of the United States were tracted in 1970, particularly small rural areas, so for the minority of respondents who lack census tract identifiers, we use minor civil divisions (MCDs) to define neighborhood, which are somewhat larger neighborhood constructs. Although these families may not have been social neighbors in the sense of interacting closely with each other, they did live in close geographic proximity to each other and this neighborhood construct should capture

important environmental influences. Census tracts typically comprise approximately 5000 people, and due to the cluster design, respondents in urban areas may have lived just a city block apart. In rural areas, the families were spread farther apart, but still were among each other's closest neighbors due to the cluster design. The PSID cluster design is discussed in greater detail in Solon et al. (2000).

Two samples are examined, one each to estimate the sibling and neighbor correlations. There are three criteria for inclusion in the sibling sample: (1) the individual was 0-16 years old in 1968 residing in an original PSID sample family, (2) the individual has a valid measure for the health indicator under consideration[3], and (3) the individual has at least one sibling who also meets these two criteria. For the neighbor sample, individuals must meet criteria (1) and (2) above, and individuals must have at least one unrelated childhood neighbor who also meets these two criteria.

Attrition of children from the PSID is not purely random; rather, it is disproportionately concentrated among children from lower-income families (Solon 1992; Fitzgerald, Gottschalk, & Moffitt 1998). In addition, the PSID over-sampled low-income households.[4] Following Hill (1992), we therefore use the most recent PSID individual population weights, corresponding with the year in which the relevant outcome variable of interest is measured, normalized per individuals in the respective analyses. The PSID population weight assigned to an individual reflects the number of individuals represented by a particular respondent, compensating for both unequal selection probabilities and differential attrition. The effective sample size is retained in these

[3] For health indicators measured in adulthood, we examine the most recent valid measure between 1986 and 2001.

[4] The overall pattern of results reported in this paper is robust to the exclusion of the PSID Survey of Economic Opportunity (SEO) "poverty" sub-sample.

analyses with the use of a normalized weight, while generating unbiased population estimates for generalizing to a national population in the U.S. in this age range.

The PSID is the only survey that contains information on adult health and health behavior, the neighborhood in which the person grew up, along with the same information on that individual's siblings and childhood neighbors. In addition, the PSID over-sampled minority and low-income families, which generate sufficient medium to high poverty neighborhoods that allow investigation of nonlinear neighborhood effects. That is, although the analyses are not included in the current draft, we will investigate whether neighborhood effects are linear or emerge only at some threshold, such as high poverty concentrations. Similarly, we will also examine differential impacts of neighborhood effects by race/ethnicity and parental education.

The sample sizes differ depending on the health indicator being examined (Table 1). For the sibling sample, the largest (smallest) sample is for low birth weight (self-assessed health in childhood for males) and consists of 3,678 (565) individuals in 1,202 (249) families. For the neighborhood sample, the largest (smallest) sample is for low birth weight (cigarette smoking in adulthood for males) and consists of 2,258 (459) individuals in 881 (307) families in 327 (127) neighborhoods.

HEALTH OUTCOMES

The empirical analysis examines five dimensions of health, which are discussed in turn. The neighbor sample means for the various health-related outcomes are displayed in Table 1a.

Low Birth Weight. Low birth weight is the single-most important predictor of infant mortality, which remains a significant public health problem in the U.S., especially

among poor and minority populations. For example, decreasing the incidence of low birth weight through increases in the use of prenatal care was the primary motivation for the changes in the Medicaid coverage of pregnant women that took place during the 1980s (Currie & Gruber 1996). Low birth weight is also an important outcome because it is associated with significant risks in children's health and cognitive development (Conley & Bennett 2000; McCormick, Brooks-Gunn, & Workman-Daniels 1992; Boardman et al. 2002), and health problems later in life, such as coronary heart disease in adulthood (Barker 1995). Low birth weight could thus lead to health inequalities in childhood and beyond.

Previous research on birth outcomes has focused on individual-level determinants and documented the importance of maternal age, marital status, income, education, and the timing of prenatal care. However, these individual-level determinants collectively explain only a small proportion of the total variability in birth weight and infant mortality (Ellen et al. 2001).

Birth weight is well-suited for studying the effects of neighborhood context because it is sensitive to short-run influences on maternal health during the length of pregnancy (Morenoff 2002). In our analysis, we take the standard approach of defining low birth weight as birth weight of less than 2500 grams (5 pounds 8 ounces). Because low birth weight is a rare event (it occurs in six percent of the births in this sample), it is difficult in our multilevel framework to detect variation due to the small within-neighborhood sample sizes (each neighborhood contains on average 8 observations). In spite of these difficulties, our results reveal significant neighborhood effects (as will be discussed in section VI).

Self-assessed Health. Our primary health measure in childhood and adulthood is individuals' self-reported health status. For health status in years of adulthood, each respondent was asked: "How would you describe your health generally? Would you say it is excellent, very good, good, fair, or poor?" For individuals with multiple self-reports in adulthood, we use the most recent self-report of health status.

Health status in childhood was asked retrospectively in a similar way, "Would you describe your health in childhood (i.e., between the ages 0-16) as excellent, very good, good, fair, or poor?" This question was added to the 1999 and 2001 PSID surveys. The retrospective nature of the childhood health status responses is likely to induce some measurement error, likely reducing the estimated correlations.

Self-reported health has been shown to be closely linked to morbidity reported in surveys or diagnosed through clinical examinations (Larue et al. 1979; Linn et al. 1980; Mays et al. 1992). It is also one of the most powerful predictors of mortality, even when controlling for physician-assessed health status and health-related behaviors, and it is a strong determinant of whether patients choose to use medication and health services. In our analysis we use a dichotomous measure of problematic health status because the lower tail of the distribution of self-assessed health is both most persistent across adjacent years (minimizing potential biases due to measurement error) and most strongly predictive of morbidity and mortality. In adulthood, the indicator is equal to 1 if individuals responded that their health was either fair or poor, and equal to 0 if their health was excellent, very good, or good. In childhood, the indicator takes on values of 1 if individuals responded that their health in childhood was good, fair, or poor, and 0 otherwise. The average age corresponding to our measure of health in adulthood is 39.

About seven percent of our sample had problematic health (as defined above) in adulthood.

Obesity in Early to Mid Adulthood. Neighborhood conditions, such as the availability of recreational facilities and access to low cost healthy foods, may make it more or less costly to undertake health-promoting behavior, such as exercising regularly and eating nutritious foods (Robert & Yen 1998). These health-behavior habits are formed to a large degree in childhood and adolescence. Poorly maintained neighborhood environments may manifest themselves in crumbling sidewalks and dangerous playgrounds, and act to undermine health-promoting efforts in youth. Previous studies have shown that African-American neighborhoods are more likely to suffer from institutional risk factors such as the proliferation of liquor stores, insufficient supplies of nutritious foods in local grocery stores, and insufficient supplies of prescription drugs at local pharmacies (Morenoff & Lynch 2002; LaVeist & Wallace 2000).

In this paper, we use weight and height information measured in adulthood to create body mass index (BMI) for each individual in our sample.[5] In our analyses, we estimate sibling correlations and neighbor correlations in obesity to investigate the relative importance of contextual-level factors. We conduct separate analyses by gender, and we use the standard convention of defining obesity as BMI of greater than or equal to 30. The average age corresponding to our measure of BMI in adulthood is 39. Twenty-one percent of our sample is obese.

Health Behaviors—Smoking. Almost 50% of deaths in the U.S. are preventable and directly attributable to behavioral causes. Despite the dramatic increase in the

[5] The national adult distribution of BMI estimated using the PSID, which is based on telephone reports of height and weight, align quite closely with the distribution estimated from the National Health and Nutrition Examination Study, which obtains clinical assessments of height and weight (Stafford, 2002).

public's awareness of the health hazards of smoking that has occurred over the past two decades, cigarette smoking and consumption of alcohol cost upward of 100 billion dollars annually in health care costs and lost productivity (Manning et al. 1991).

One of the strongest predictors of adult substance use problems is early onset of substance use (Reardon et al. 2001; Kandel et al. 1992). Among individuals who have ever smoked cigarettes regularly, two-thirds began smoking before age 19 (based on the PSID). Previous research on early cigarette use has focused on individual- and family-level factors, documenting a significant relationship between family socioeconomic status and young people's likelihood of smoking. There are differing views regarding whether these health behavior patterns reflect a health lifestyle orientation or are responses to behavioral incentives resulting from neighborhood contextual-level pressures.

Our analysis is the first to investigate both the effects of neighborhood context on the age of initiation of cigarette smoking in adolescence, and the effects of the neighborhood of upbringing on the subsequent smoking behavior of these same individuals in adulthood.[6] Roughly half of both males and females in our sample had ever smoked cigarettes. The average age corresponding to our measure of smoking behavior in adulthood is 40. Twenty-seven percent of males and 22% of females in our sample currently smoke in adulthood.

Risk Preferences. As described in Section II, risk preferences may influence health behaviors such as smoking. More generally, risk preference parameters shape a wide array of behavioral choices of individuals, and thus analysis of risk preferences and

[6] In the 1999 & 2001 PSID surveys, respondents were asked, "Do you currently smoke cigarettes?" "Have you ever smoked?" If they answered yes to either question, they were then asked "About how old were you when you first began smoking cigarettes/first smoked a cigarette?"

how they are shaped by childhood neighborhood influences has far-reaching implications and can contribute to our understanding of many different phenomena.

We use data from a 1996 supplement to the PSID to explore the extent to which an index of risk tolerance measured in adulthood is correlated between siblings and correlated between childhood neighbors. Our measure of risk tolerance is developed from a series of questions asked of respondents about the circumstances under which they would take different hypothetical gambles. The PSID risk tolerance measures are computed from an identical set of questions to those used by Barsky et al. (1997), who show that these measures predict risky behaviors, including smoking, by respondents in the Health and Retirement Survey (HRS).

The questions in the PSID are as follows: "Suppose you had a job that guaranteed you income for life equal to your current, total income. And that job was (your/your family's) only source of income. Then you are given the opportunity to take a new, and equally good job with a 50-50 chance that it will cut your income by one-third, or, on the other hand, it could double your income with a 50-50 probability. Would you take that new job?" Based on the response to that question, the PSID asks follow-ups about jobs that double their income with a 50 percent probability or either cut your income by 10%, 20%, 50%, or 75% with a 50 percent probability. The risk aversion questions were only asked of 1996 PSID household heads who were working

Assuming a CES utility function and correcting for measurement error, PSID respondents can be sorted into four distinct levels of risk tolerance (high risk tolerance, medium risk tolerance, low risk tolerance, and very low risk tolerance), based on their responses to these questions. Barsky et al. (1997) summarize the procedure on how the

risk aversion parameters are computed using the HRS data. The same procedure was used to compute the risk aversion measures using the PSID data (Luoh & Stafford, 2001). Assuming CES preferences, the four categories (high risk tolerance, medium risk tolerance, low risk tolerance, and very low risk tolerance) correspond to estimated risk aversion measures of 1.75, 2.86, 3.57, and 6.67, respectively (Barsky et al. 1997). Our estimates indicate that roughly one-quarter of males in our sample have high risk tolerance.

V. ECONOMETRIC MODEL

We begin by assuming the true model for the health outcome of interest is:

$$H_{nfs} = \alpha' X_{nf} + \beta' Z_n + \varepsilon_{nfs} \qquad (1)$$

where H_{nfs} denotes health status for sibling *s* in family *f* in neighborhood *n*, X_{nf} is the vector that includes all family characteristics (measured and unmeasured) that affect H_{nfs}, Z_n is the vector of all neighborhood characteristics that affect H_{nfs}, and ε_{nfs} is the error term that includes all individual-specific factors that are not related to X_{nf} or Z_n. Note that for illustrative simplicity, at this juncture, we do not attempt to incorporate potential interactions between family and neighborhood background effects or nonlinearaties into the model, but rather assume a linear representation.

Due to the self-selection of advantaged families sorting into advantaged neighborhoods for the reasons discussed in section II, we expect the family background effect, X_{nf}, and the neighborhood background effect, Z_n, to be positively correlated. Because it is difficult to fully and accurately measure every factor in X_{nf} and Z_n, the assumption that ε_{nfs} is uncorrelated with the observable measures of X_{nf} and Z_n, will be

violated, leading to biased estimates of neighborhood effects (β) and family background effects (α). Using the taxonomy of Manski (1993), it is not possible to distinguish the two types of "social effects" ("endogenous effects" and "exogenous effects") from the nonsocial "correlated effects". Manski also demonstrates it is not possible to distinguish the two types of social effects from each other.

Therefore, the primary goal of our analyses is focused on an overall assessment of the relative contributions of individual, family, and neighborhood effects on health-related outcomes. We also analyze the relative contribution of a parsimonious set of measured individual, household, and neighborhood covariates to the total variation from each component, and test specific hypotheses about the effects of specific characteristics of households and neighborhoods.

Our strategy for assessing the importance of contextual effects involves estimating the fraction of variation in health outcomes of interest that lies between families and neighborhoods, to provide an upper bound on the possible effect of these contexts. The intuition motivating the use of this strategy is that if family background and residential community are important determinants of health outcomes, there will be a strong correlation between siblings in their health outcomes, as compared to two arbitrarily chosen individuals. And if the neighborhood where the child grew up is important, it will show up as a strong correlation between neighboring children's subsequent health outcomes.

As demonstrated in Solon et al. (2000), using the additive model of the effect of family and neighborhood context in equation (1), the population variance of H_{nfs} can be decomposed as:

$$Var(H_{nfs}) = Var(\alpha' X_{nf}) + Var(\beta' Z_n) + 2Cov(\alpha' X_{nf,} \beta' Z_n) + Var(\varepsilon_{nfs}). \quad (2)$$

Similarly, the covariance in H_{nfs} between siblings s and s' is:

$$Cov(H_{nfs}, H_{nfs'}) = Var(\alpha' X_{nf}) + Var(\beta' Z_n) + 2Cov(\alpha' X_{nf,} \beta' Z_n). \quad (3)$$

The sibling correlation, $\text{cov}(H_{nfs}, H_{nfs'}) / \text{var}(H_{nfs})$, measures the proportion of the total variation in the health outcome under consideration due to factors shared by siblings. From (3) we see that siblings have correlated health outcomes because they have shared family and neighborhood backgrounds, corresponding to the first and second terms of (3), respectively. The sorting of families into neighborhoods is reflected in the third term. The sibling covariance then captures all measured and unmeasured factors shared by siblings that may have an impact on health outcomes, such as the socioeconomic status of parents, genetic traits shared by siblings, family structure, as well as neighborhood effects stemming from the quality of neighborhood conditions.

Augmenting the estimation of sibling correlations with the estimation of neighbor correlations enables us to bound the relative importance of family and neighborhood factors. To see this, note the covariance between neighbors is:

$$Cov(H_{nfs}, H_{nf's'}) = Cov(\alpha' X_{nf}, \alpha' X_{nf'}) + Var(\beta' Z_n) + 2Cov(\alpha' X_{nf,} \beta' Z_n). \quad (4)$$

The last two terms in (3) and (4) are identical, so we expect the covariance between neighbors to be smaller than the covariance between siblings because siblings share both the neighborhood and the same family. As Solon et al (2000) state, if the covariance among neighbors is small relative to the covariance among siblings, the family effects, which are represented by the first term in (3), must be the main source of the covariance among siblings. Previous studies of sibling correlations do not disentangle family from

neighborhood effects, exceptions being Solon et al (2000, 2001), Raaum, Salvanes, and Sorensen (2002), and Oreopoulos (2002), none of whom examine health outcomes.

The neighborhood correlation, $\text{cov}(H_{nfs}, H_{nf's'}) / \text{var}(H_{nfs})$, measures the proportion of the variation in the health outcome that can be attributed to factors shared by individuals from the same neighborhood. In (4), we notice that the neighborhood covariance consists of more than the variance in neighborhood characteristics given in the second term, and it should therefore be viewed as an upper bound on the neighborhood influence on the covariance in H_{nfs} between neighbors. The first and third terms are both expected to be positive, leading to an upward bias. The first term represents the sorting of similar families into the same neighborhoods, since neighboring children share similar family characteristics. Similarly, the third term also represents sorting, in that it captures sorting of (dis)advantaged families into (dis)advantaged neighborhoods. We see that positive sorting, $Cov(\alpha' X_{nf}, \alpha' X_{nf'}) \geq 0$ and $Cov(\alpha' X_{nf,} \beta' Z_n) \geq 0$, implies that $Var(\beta' Z_n) \leq Cov(H_{nfs}, H_{nf's'})$.

Access to neighborhood identifiers and family characteristics in the same data enables us to tighten the upper bound on the neighborhood effect and also establish a lower bound on the family effects. First, it follows from (4) that the upper bound on the neighborhood effects can be made tighter by introducing observable family characteristics shared by the neighbors, and by subtracting that as an observable part of the first term of (4). Following Solon et al. (2000) and Altonji (1988), we estimate the part of $\alpha' X_{nf}$ related to observable family characteristics such as parental income, education, family structure, and race. Let $\widetilde{X}_{nf}$ denote the observable subset of family

characteristics with associated parameters $\hat{a}$ estimated *within* neighborhoods. We can then subtract off the sorting component arising from the fact that similar families tend to cluster in neighborhoods,

$$Cov_{adj}(H_{nfs}, H_{nf's'}) = Cov(H_{nfs}, H_{nf's'}) - Cov(\hat{\alpha}'\widetilde{X}_{nf}, \hat{\alpha}'\widetilde{X}_{nf'}) \ . \qquad (5)$$

This approach estimates only the *direct* effects of neighborhood on health outcomes. If neighborhood factors caused parents to obtain higher paying jobs, for example, that effect would be attributable to the family component and not the neighborhood; this indirect effect would not be captured by (5).

Second, the sibling correlation can be decomposed into a part arising from to shared neighborhood origins and a part related to family background characteristics. Specifically, the difference between the sibling correlation and the adjusted neighbor correlation represents a lower bound of the magnitude of the (composite) effect of family background on the health outcome of interest.

LIMITATIONS OF THE APPROACH

There are several factors that could cause us to underestimate the magnitude of neighborhood effects by employing the strategy of neighbor correlations. First, because siblings share similar family environments for longer periods than neighboring children share neighborhood environments, we expect lower correlation for neighbors for health outcomes that can be viewed as a product of long-term process. We estimate the correlation between individuals that were childhood neighbors in 1968, but if 1968 neighborhood is a poor proxy for longer-run neighborhood environment, our estimates of the influence of neighborhoods may be subject to a downward errors-in-variables bias. The potential for measurement error is a serious concern since residential mobility is

common in the United States, especially among families with younger children. Thus, children sharing a neighborhood at any given point in time may have quite different residential histories. However, Kunz et al. (2002) investigate this issue using the PSID and show the high degree of persistence in the quality of children's neighborhood environments. Kunz et al. (2002) estimate the autocorrelations of observed neighborhood characteristics inhabited by the PSID children, and find the sample autocorrelation between the average of log mean income during the 1970-1980 period and each single year value is at least 0.90 for every year and averages 0.94. This evidence minimizes concerns about this type of potential bias, since residential moves typically occur between neighborhoods of similar quality. In future analyses we will also examine the robustness of the correlation estimates to the duration of residence in the child's 1968 neighborhood; the neighborhood effect should be higher for children who have remained in the same community for a longer duration. However, Solon et al (2000) find that neighbor correlations in education are not sensitive to restrictions to neighbors who have been long-time residents of the same neighborhood, implying that our current estimates may be robust to this factor.

A second factor that could cause us to underestimate the magnitude or importance of neighborhood effects is that there is no easy way with the correlation method of allowing nonlinear neighborhood effects, so this technique is limited in assessing how living in an extremely disadvantaged (advantaged) neighborhood affects health-related outcomes. For example, neighborhood problems such as violence, lead paint, or pollution, may only have significant impacts on health when they achieve some threshold of incidence. Although it is not contained in this draft, we plan to investigate whether

neighborhood effects are linear or emerge only at some threshold, such as high poverty concentrations. In addition, certain groups may exhibit particular sensitivity to the quality of neighborhood conditions. For example, families who lack social and economic resources may be more vulnerable to poor neighborhood conditions. Therefore, in the future we will examine differential neighborhood correlations by race/ethnicity and socioeconomic status.

Finally, and important from a policy perspective, effect sizes that program evaluators commonly view as medium or even large may translate into small proportions of variance explained by neighborhood background and into small intraneighborhood correlations (Duncan & Raudenbush (2001); Cain & Watts 1972; Rosenthal & Rubin 1982). As highlighted by Duncan & Raudenbush (2001), a small correlation between neighbors does not rule out a large effect size associated with a measured difference between neighborhoods. To illustrate this point, they consider a standardized mean difference between a set of experimental neighborhoods and an equal number of control neighborhoods, and show that even a standardized effect size of d=.80, which is commonly viewed as very large, translates into an intraneighborhood correlation of .14.[7]

V. ESTIMATION METHODS

To motivate the estimation methodology, first consider the decomposition of variance in the following simple two-level random-intercept model:

[7] Using the hypothetical example discussed in Duncan & Raudenbush (2001), assume we have a treatment group of neighborhoods and control group of neighborhoods, and we compute the standardized mean difference, *d,* between these two groups of neighborhoods. The proportion of variance explained by treatment is then *R*-squared, which in this case is $R^2 = d^2/[\, d^2 + 1/p(1\text{-}p)]$, where p is the proportion in the treatment group and $(1\text{-}p)$ is the proportion in the control group. In this hypothetical example, all between-neighborhood variance is created by the treatment. Thus, *R*-squared is equivalent to the intra-neighborhood correlation. In non-experimental settings, we do not have treatment groups, but rather sets of "high-risk" and "low-risk" neighborhoods.

$$H^*_{fs} = \mu + \phi_f + \varepsilon_{fs} \ ,$$

where *f* indexes families and *s* indexes individuals. Here, the total variance of H^*_{fs} can be decomposed as:

$$\text{var}(H^*_{fs}) = E[\text{var}(H^*_{fs} \mid \phi_f)] + \text{var}[E(H^*_{fs} \mid \phi_f)] .$$

The first term of the decomposition is the intrafamily variance—i.e., the part of the variance that is not due to the variability of ϕ_f (the family effect). The second term, $\text{var}[E(H^*_{fs} \mid \phi_f)]$, is the inter-family variance—i.e., the part of the $\text{var}(H^*_{fs})$ due to the variability of ϕ_f (the family effect). Here, the intrafamily correlation is the fraction of variance that lies between families, $\frac{\text{var}[E(H^*_{fs} \mid \phi_f)]}{\text{var}(H^*_{fs})}$, and is the relevant part of the variance that is due to the heterogeneity of means between families.

In order to decompose the total variation in the health outcome of interest into the fraction that lies between neighborhoods, families, and individuals, we estimate a three-level hierarchical random effects model. Our data are hierarchical because we have data on individuals who are nested within families, which are nested within neighborhoods. Multilevel modeling techniques can accommodate the hierarchical and unbalanced structure of our data, non-independence of the (sometimes overlapping) pairs of siblings and neighbors, as well as the non-normality of our health outcomes of interest (Raudenbush & Bryk 2002).

All of the health outcomes that will be analyzed in this paper are binary outcomes in which the binary response is interpreted as the result of an underlying latent process. The three-level hierarchical random effects model for our binary health outcomes can be

derived through a latent variable conceptualization. Specifically, we assume that there exists a latent continuous variable H^*_{nfs} underlying H_{nfs}. We observe only our binary health outcome H_{nfs} directly, but not H^*_{nfs}. We know, however, $H^*_{nfs} > 0$ if $H_{nfs} = 1$, and $H^*_{nfs} \leq 0$ if $H_{nfs} = 0$. For example, H^*_{nfs} may represent a continuous scale of health status, but we observe whether the individual is in excellent/very good/good vs. fair/poor heath.

We estimate the three-level hierarchical random effects model given by

$$H^*_{nfs} = \beta' X_{nfs} + \eta_n + \phi_{nf} + \varepsilon_{nfs} \qquad (6)$$

where β is a vector of regression coefficients corresponding to the effect of covariates X_{nfs} (which represent observed characteristics of the neighborhood, the family, and the individual), and η_n and ϕ_{nf} are the random effects, which represent unobserved characteristics of the neighborhood and the family, respectively.[8] In this formulation the random effects, which play the role of additional error terms, are assumed to be normally distributed with mean 0, and var(η_n)=σ^2_n and var(ϕ_{nf})=σ^2_f. Here ε_{nfs} is an individual error term associated with individual *s* from family *f* in neighborhood *n* and is assumed to have a standard logistic distribution with mean 0 and variance $\frac{\pi^2}{3}$ (where $\pi \approx 3.14$).

In this model, individuals from the same neighborhood but not in the same family (i.e., neighbors) are correlated because they share the random effect η_n, and siblings are correlated because they share the random effects η_n and ϕ_{nf}. In this model, the sibling correlation and neighbor correlation can be computed, respectively, as:

[8] Maximum-likelihood (ML) estimates based on a numerical integration procedure were computed using the gllamm6 macro in Stata (Rabe-Hesketh et al. 2000).

$$\rho_{\text{sibling}} = \frac{\sigma_n^2 + \sigma_f^2}{\sigma_n^2 + \sigma_f^2 + \frac{\pi^2}{3}} \quad ; \quad \rho_{\text{neighbor}} = \frac{\sigma_n^2}{\sigma_n^2 + \sigma_f^2 + \frac{\pi^2}{3}} \quad .$$

The sibling correlation is between the unobserved latent variables H^*_{nfs} & $H^*_{nfs'}$; the neighbor correlation is between the unobserved latent variables H^*_{nfs} & $H^*_{nf's'}$.

Our health-related outcomes vary with age and gender. Because we did not want our correlations to reflect the influence of either of these two demographic factors, we adjusted for them in our baseline model by including age as an explanatory variable in the vector X_{nfs}, and conducting separate analyses by gender. An exception is birth weight, where limited sample sizes required us to pool boys and girls and control for sex. Moreover, given that age affects health outcomes and that most same-aged children do not belong to the same family, it is important to control for age in the baseline model. Otherwise, between-family variance could mostly reflect differences between individuals of different ages. Separate analyses by gender allow sibling correlations to differ between brother-brother pairs, sister-sister pairs, and brother-sister pairs; and they allow neighboring boys to have a different correlation in health outcomes than neighboring girls (i.e., allow childhood contextual-level effects on health-related outcomes to differ by gender).

We then estimate "adjusted neighbor correlations," which are net of the similarity arising from childhood neighbors having similar family background characteristics. To extract the impact of similar family backgrounds out of the neighbor correlation, we first estimate the following regression:

$$H^{*}_{nfs} = \alpha_1(fa\min c_{nf} - \overline{fa\min c_{n\bullet}}) + \alpha_2(black_{nf} - \overline{black_{n\bullet}}) + \alpha_3(femhd_{nf} - \overline{femhd_{n\bullet}}) + \varepsilon_{nfs},$$

(7)

where $faminc_{nf}$ is log family income in 1967 (as reported in 1968); $\overline{fa\min c_{n\bullet}}$ is the neighborhood mean of log family income (based on 1970 Census data); $black_{nf}$ is a dummy variable equal to one if the 1968 head of the household was black; $\overline{black_{n\bullet}}$ is the fraction of individuals in the neighborhood that are black (based on 1970 Census data); $femhd_{nf}$ is a dummy variable equal to one if the 1968 head of the household was female; and $\overline{femhd_{n\bullet}}$ is the fraction of households with children that are headed by females in the neighborhood (based on 1970 Census data). The neighborhood means of family income, racial composition, and family structure were obtained by merging on aggregated 1970 Census data at the census tract level. Using the within-neighborhood estimates of the family background effects of parental income, race, and family structure on the relevant health outcome, will ensure the coefficients (α) will not be biased by omitted neighborhood variables. In combination, the resulting estimates of the effects of these family background characteristics can be taken as a conservative estimate of $\alpha' X_{nf}$ in equation (1).

We then estimate the inter-neighbor variance in $\hat{\alpha}' X_{nf}$ by estimating a hierarchical random-intercept model of $\hat{\alpha}' X_{nf}$ on a neighborhood-level random effect and a family-level random effect. We then subtract our estimate of the inter-neighbor variance in $\hat{\alpha}' X_{nf}$ from the estimate of the overall inter-neighbor variance in H^{*}_{nfs}.

Dividing the resulting quantity by $\hat{Var}(H^*_{nfs})$ yields a tighter upper bound on the proportion of $Var(H^*_{nfs})$ that can be attributed to neighborhood effects.

AGE OF INIATIATION OF SMOKING

For smoking initiation we also estimate a three-level hierarchical random effects discrete-time hazard model to analyze the age of onset of cigarette use in adolescence. The hazard function, h_{nfst}, is the probability that individual *s* from family *f* in neighborhood *n* begins smoking cigarettes in year *t,* given the individual has never smoked cigarettes in any previous year. The hazard is specified in a logit form, where in the baseline model, the explanatory variables include only a set of age dummy variables (AGE_{nfst}) and a neighborhood-level random effect (η_n) and family-level random effect (ϕ_{nf}):

$$h_{nfst} = 1/(1-\exp[\sum_{t=10}^{18}\alpha_t(AGE_{nfst})+\eta_n+\phi_{nf}]) \quad .$$

In this model, we are implicitly assuming proportional odds—in particular, we assume the baseline logit hazard curves in the *J* neighborhoods are parallel to one another, and the baseline logit hazard curves in the *K* families in these neighborhoods are parallel to one another. We, however, conduct the analyses separately by gender to allow the relative importance of neighborhood and family contextual-level influences to differ for boys and girls.

VI. RESULTS

We first present the unadjusted sibling and neighbor correlations in each of the health-related outcomes, and examine how much of these effects can be explained by the fact that families in a neighborhood tend to be similar. We then conduct a series of

simulations to attempt to shed light on what these correlation estimates imply in terms of the absolute size of the effects of family and neighborhood background. We also perform Blinder-Oaxaca decompositions to assess the relative roles of family background and neighborhood quality during childhood in contributing to racial health disparities. In particular, we simulate the effects of giving blacks the distribution of family background and neighborhood quality during childhood that whites possess, respectively, to examine how much of the black-white gap in early and mid life health status can be explained by these factors.

SIBLING AND NEIGHBOR CORRELATIONS

The sibling and neighbor correlation estimates are based on the decomposition of variance into the fraction that lies between neighborhoods, families, and individuals, for the relevant health outcome. The estimates for each of the health-related outcomes are reported in Table 2, and discussed in turn.

For most health status outcomes, sibling correlations are high. Siblings' birth weight demonstrates the highest correlation at 0.63. Sibling correlations in self-assessed health in adulthood and obesity in adulthood differ slightly between brothers and sisters, but hover around 0.50. Measurement error in self-reported health in childhood due to the retrospective nature of respondent's reports is likely producing a downward bias in the estimated childhood health status correlations.

Health behaviors, at least smoking, also demonstrate high correlations within families. The correlation is 0.452 for sisters and higher yet at 0.56 for brothers. Risk tolerance, which can only be estimated for males because the question was asked of too

few females, has a sibling correlation of 0.384. In other words, knowing a brother's risk tolerance in adulthood helps predict 38 percent of another brother's risk tolerance.

As discussed in section V, sibling correlations by themselves cannot disentangle how much of the resemblance among siblings in their health outcomes is due to the effects of family background and how much is due to the effects of neighborhood quality during childhood. Augmenting the sibling correlation estimates with corresponding neighbor correlation estimates, reveals family background is the most important determinant for each of the health outcomes. The neighbor correlation for each of the health outcomes is considerably smaller than the corresponding value for siblings, but is still significant. For example, the results indicate that knowing a childhood neighbor's health status at birth or knowing a childhood neighbor's health status during childhood predicts roughly 10 percent of another childhood neighbor's health status at birth and during childhood, respectively. Interestingly, the neighbor correlation estimates for these health outcomes in early life are similar in magnitude to those reported by Solon et al. (2000) for educational attainment.

We see striking gender differences in health status in adulthood with males exhibiting substantially larger neighbor correlations than females. In particular, the estimated neighbor correlation in adulthood health status among males is 0.257, while 0.115 among females. In other words, knowing the adulthood health status of a male childhood neighbor predicts about 25 percent of the adult health status of another male childhood neighbor. These gender differences suggest the quality of neighborhood conditions during childhood have a much larger impact on males' subsequent health in adulthood. Furthermore, by comparing the magnitudes of the sibling and neighbor

correlations in adulthood health among males, the results indicate that more than half of the 0.452 sibling correlation is attributable to neighborhood effects. From the adjusted neighbor correlation estimates, we also find that observable family sorting does not seem to explain all the resemblance in adulthood health status among persons who grew up in the same neighborhood. Across all health measures, the adjusted neighbor correlation is 10-20% lower than the unadjusted neighbor correlation. In future work we will consider a broader array of family factors. Taken together, these gender differences suggest that while family background is the primary gatekeeper of the intergenerational transmission of adulthood health status among females, neighborhood quality during childhood is a significant gatekeeper of the intergenerational transmission of adulthood health status among males.

A potential explanation for the gender differences in the relative roles of family background and neighborhood quality during childhood lies in considering gender differences in the processes that lead to the intergenerational transmission of economic status. In particular, the father's socioeconomic status is the primary driver of the intergenerational transmission of economic status among males (Solon 199?). As a result, there is likely a significant correlation between the neighborhood quality an individual is exposed to in childhood and in adulthood (due to patterns of economic residential segregation). That is, there is a significant degree of persistence in exposure to neighborhood quality over the life course (we plan to explore this in future research). On the other hand, females' intergenerational socioeconomic mobility is much closer linked to marriage because males are usually the breadwinner of the family. Thus, females may be expected to exhibit lower correlations between neighborhood quality in

childhood and in adulthood (as long as assortative mating along socioeconomic status is not extreme). These gender differences in the processes of intergenerational economic mobility have implications for gender differences in the processes that lead to the intergenerational transmission of health status. (develop this further and incorporate into theoretical model—cite Case et al. (2002,2003)).

As individuals age, the total amount of lifetime exposure to social factors that may influence their health status increases. This pattern may explain the relatively high sibling correlations in birth weight that are attributable to family background effects (0.55) versus sibling correlations in adulthood health status that are attributable to family background effects (0.21 among males). That is, social factors can influence birth weight only through their influence on the mother while the child is in uterine. But during their childhood and adulthood, individuals themselves are continuously exposed to various social factors, and these factors likely differ among siblings, especially once they leave their parents' home

The estimated neighbor correlation for obesity among females is about 0.1 and suggests that while obesity is driven by hereditary risk factors and family background influences on health lifestyle orientation, neighborhood factors exert a significant effect on the prevalence of obesity. In particular, knowing the adult obesity status of a female childhood neighbor predicts about 10 percent of the adult obesity status of another female childhood neighbor. The neighbor correlation in obesity for males is roughly half the magnitude of females. Females' exercise and healthy eating behaviors (dietary preferences), which are formed in large part in childhood and adolescence, may be more sensitive to neighborhood resources such as the availability of youth recreational

facilities and/or supplies of nutritious foods in local grocery stores. The greater impact of neighborhoods on obesity among females may also be the result of peer effects that produce a greater value on physical appearance in terms of self-esteem among girls.

The gender differences in neighbor correlations are most striking for cigarette smoking behavior. As discussed in section V, we begin by estimating a three-level (hierarchical) random effects discrete-time hazard model to analyze the age of onset of cigarette use in adolescence, separately for boys and girls. Roughly 30 percent of both boys and girls in the sample smoked cigarettes in adolescence. The results are presented in Table 3. Estimates of the random effects of the neighborhood (σ_n) and family (σ_f) components indicate that, for boys, neighborhood and family background have very large and significant effects on the likelihood of youth smoking initiation.[9] Interestingly, for boys, the neighborhood and family effects are of similar magnitude. For example, the estimated σ_n of 0.9154 implies the odds of adolescent smoking initiation for boys who grow up in neighborhoods that are one standard deviation below average neighborhood quality are 2.5 times [exp(.9154)] the corresponding odds of individuals who grow up in neighborhoods of average quality. In contrast, for girls, while family background has very large and significant effects, neighborhood effects on adolescent smoking initiation are half the magnitude of family effects and are only marginally significant.

We next examine the estimated sibling and childhood neighbor correlations in smoking behavior in adulthood. We find that the gender differences in the relative roles of family background and neighborhood quality on smoking behavior in adulthood mirror those found in adolescence. As shown in the fifth column of Table 2, the estimated

[9] Likelihood ratio tests are used to test the statistical significance of the family and neighborhood random effects.

neighbor correlation in adulthood smoking behavior for males is 0.27, while the neighbor correlation for females is negligible. In other words, knowing the adult smoking behavior of a male childhood neighbor predicts 27 percent of the adult smoking behavior of another male childhood neighbor. Moreover, from the adjusted neighbor correlation estimate, we also find that observable family sorting does not seem to explain the resemblance in adulthood smoking behavior among persons who grew up in the same neighborhood. These results underscore the importance of neighborhood context in shaping the early formation of addictive health behaviors that persist throughout adulthood for males. In contrast, family background is the dominant determinant of smoking behavior in adolescence and adulthood among females.

Consistent with the above evidence on the importance of neighborhood effects on smoking behavior among males, we also find the childhood neighbor correlation in risk tolerance in adulthood is 0.183—almost half the sibling correlation (see column 6 of Table 2). The results indicate significant childhood neighbor correlations in risk preferences in adulthood, and suggest these preference parameters are shaped by childhood neighborhood influences. Risk preference parameters shape a wide array of behavioral choices of individuals (including smoking), and thus, this evidence on neighbor correlations in risk preferences has far-reaching implications and can contribute to our understanding of many different outcomes/phenomena.

MAGNITUDE OF EFFECTS OF FAMILY AND NEIGHBORHOOD BACKGROUND

What do these correlation estimates mean in terms of the absolute size of the effects of family and neighborhood background? In Table 2a, we present estimates of the standard deviation of family and neighborhood effects, respectively, for our health

outcomes of interest. We see that although the neighbor correlations may seem low, estimates of σ_n indicate that neighborhood quality have very large and significant effects on nearly all of the health-related outcomes, particularly for males. For example, the estimated σ_n of 1.399 for problematic health in adulthood among males implies that the odds of having problematic health in early to mid adulthood for males who grow up in neighborhoods that are one standard deviation below average neighborhood quality are 4 times [exp(1.399)] the corresponding odds of males who grow up in neighborhoods of average quality.

To provide further insight and facilitate interpretation of our results, we present graphically in Figures 1 and 2 the predicted probability of having problematic health at age 40 among males over the entire range of neighborhood quality and family background effects, respectively. The graphical representation of the results highlights the dramatic impact of the magnitude of the family and neighborhood effects.[10]

Specifically, Figure 1 shows the predicted probability of having problematic health at age 40 over the entire range of neighborhood quality effects, for males with average family background, one standard deviation below average, and one standard deviation above average family background. The differences are especially striking between below average and average families in the probabilities of having problematic health in mid adulthood.

Similarly, Figure 2 shows the predicted probability of having problematic health at age 40 over the entire range of family background effects, for males who grew up in neighborhoods of average quality, one standard deviation below average, and one

[10] In future work, we plan to also explore potential interactions between family and neighborhood effects—e.g., families who lack social and economic resources may be more vulnerable to poor neighborhood conditions (Currie & Hyson 1999).

standard deviation above average neighborhood quality. The differences are especially striking between below average and average neighborhoods in the probabilities of having problematic health in mid adulthood.

In Figures 3 and 4, we similarly present graphically the corresponding probability estimates for low birth weight over the entire range of neighborhood quality and family background effects, respectively. Although we present and discuss only the simulation results for self-assessed health in early and mid life among males, we have also performed similar simulations for the other health outcome measures.

DECOMPOSITION ANALYSIS

We next employ a variation of the Blinder-Oaxaca decomposition technique to assess the relative roles of family background and neighborhood quality during childhood in contributing to racial health disparities. In particular, we simulate the effects of giving blacks the distribution of family background and neighborhood quality during childhood that whites possess, respectively, to examine how much of the black-white gap in early and mid life health status can be explained by these factors.

For the latent variable H^*_{nfs} , the standard Blinder-Oaxaca decomposition of the black-white gap in the average value of the dependent variable can be expressed as

$$\overline{H}^{W^*}_{nfs} - \overline{H}^{B^*}_{nfs} = \beta'(\overline{X}^{W}_{nfs} - \overline{X}^{B}_{nfs}) + (\overline{\eta}^{W}_{n} - \overline{\eta}^{B}_{n}) + (\overline{\phi}^{W}_{nf} - \overline{\phi}^{B}_{nf}) \quad , \qquad (8)$$

where $\overline{X}^{j}$ is a vector of average values of independent variables for race *j*, $\overline{\eta}^{j}$ is the average neighborhood random effect for race *j*, and $\overline{\phi}^{j}$ is the average family random effect for race *j*. For our estimations, age is the only variable included in *X* and the black-white difference in age is negligible in our sample, so the first quantity on the RHS of equation (8) cancels out. The second quantity on the RHS is the part of the gap due to

racial differences in average neighborhood quality during childhood, and the third quantity is the part of the gap due to racial differences in family background. In this draft of the paper, the decompositions are similar in spirit but not equivalent to Blinder-Oaxaca decompositions, because we only focus on the contextual-level portions of the gap and estimate the relative contribution of family background and neighborhood quality.

The results of the decompositions of both the black-white gap in birth weight and mid life health status among males are reported in Table 4. Blacks have nearly twice the incidences of both low birth weight and problematic health in mid adulthood. In particular, roughly 10% of blacks in our sample were born at low birth weights as compared to 6% for whites; and about 12% of black males had problematic health at mid life as compare to 6% for white males. The decomposition results indicate that for health status at birth (birth weight), racial differences in average family background account for roughly 80 percent of the black-white gap that is due to contextual-level factors; and racial differences in average neighborhood quality account for the remaining 20 percent. The portion attributable to neighborhood quality captures only the direct influences, not previous neighborhood effects that may have affected the mother's economic status, health behaviors, or the mother's birth weight (see for example, Almond & Chay (2002) and Conley & Bennett (2002) for evidence on the intergenerational transmission of low birth weight). In contrast, for mid life health status among males, the relative roles of family background and neighborhood quality during childhood are reversed. Namely, the results indicate that for mid life health status, racial differences in average neighborhood quality during childhood account for roughly 70 percent of the black-white gap that is

attributable to contextual-level factors; and racial differences in average family background account for the remaining 30 percent.

CONCLUDING REMARKS AND DIRECTION FOR FUTURE RESEARCH

In this paper, we have used correlations collected from a nationally representative longitudinal sample of siblings and neighbors to estimate upper bounds on the possible causal effects of family background and neighborhood quality on health outcomes in early and mid life. Estimates based on three-level hierarchical logistic regression models consistently show a higher scope for family (whether emanating from nature or nurture) context than for neighborhood contexts in most of the health outcomes. However, there are stark gender differences in the relative importance of family background and neighborhood context. The greater scope for neighborhood influences on males' health status in adulthood and on males' smoking behavior in adolescence and adulthood was noteworthy. Our estimates suggest that disparities in neighborhood factors account for roughly ¼ of the variation in health status among males in mid life, and contribute significantly to current racial health disparities.

The relatively large neighbor correlations among males for health status in adulthood, smoking behavior in adolescence and adulthood, and risk preferences, suggest neighborhood factors play an important role in the intergenerational transmission of health status. The evidence presented in this propels challenges future research to further our understanding of the underlying processes that produce health disparities between different racial/ethnic and socioeconomic groups. Future theoretical models and empirical analyses must incorporate and carefully model how the timing of neighborhood

exposures intersects with the natural history/trajectory of health outcomes and how neighborhood effects may vary over the life-course.

This evidence suggests further research on the effects of particular neighborhood characteristics is strongly warranted. In order to assess the policy implications of this research, we need a better understanding of the pathways through which neighborhoods and families affect health. Peer group effects, role model effects, and contextual-complementarity effects each represent distinct influences under the umbrella of neighborhood effects, and each has different policy implications. We have focused in this paper on quantifying the potential overall magnitude of family and neighborhood effects. Disentangling the causal sources of neighborhood effects is extremely difficult (Manski 1993; Moffitt 1998), but the decomposition and investigation into the mechanisms of why neighborhoods matter are an important next step and area for future research.

REFERENCES

Aaronson, Daniel. 1998. "Using Sibling Data to Estimate the Impact of Neighborhoods on Children's Educational Outcomes". *The Journal of Human Resources*. Autumn, 33(4): 915-946.

Barsky, Robert B., F. T. Juster, M. S. Kimball and M. D. Shapiro. 1997. "Preference Paramaters and Behavioral Heterogeneity: An Experimental Approach in the Health and Retirement Study". *The Quarterly Journal of Economics*. May, 112(2): 1937-1996.

Case A, Katz L. 1991. "The Company you Keep: the Effects of Family and Neighborhood on Disadvantaged Families". National Bureau of Economic Research Working Paper no. 3705.

Case, Anne, Darren Lubotsky and Christian Paxson. 2002. "Economic Status and Health in Childhood: The Origins of the Gradient". *American Economic Review*. December, 92(5): 1308-1334.

Case, Anne, Angela Fertig and Christina Paxson. 2003. "The Development of the Gradient: Earnings and Health Dynamics". Unpublished manuscript, Center for Health and Wellbeing, Princeton University.

Chandra, Amitabh and Jonathan S. Skinner. 2003. "Geography and Racial Health Disparities". Unpublished manuscript. Department of Economics and Center of Evaluative Clinical Sciences, Dartmouth Medical School.

Corcoran M, Gordon R, Laren D, Solon G. 1992. "The Association Between Men's Economic Status and their Family and Community Origins". *Journal of Human Resources.* 27: 575-601.

Crane J. 1991. "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping out and Teenage Childbearing". *American Journal of Sociology.* 96: 1226-1259.

Currie, Janet and Rosemary Hyson. 1999. "Is the Impact of Health Shocks Cushioned by Socioeconomic Status? The Case of Low Birthweight". *The American Economic Review*. May, 89(2): 245-250.

Dow, William H., J. Holmes, T. Philipson, and X. Sala-i-Martin.1995. "Disease Complementarities and the Evaluation of Public Health Interventions". Working Paper no. 5216. National Bureau of Economic Research.

Dow, William H., T. Philipson, X. Sala-i-Martin. 1999. "Longevity Complementarities under Competing Risks". *The American Economic Review*. Dec., 89(5): 1358-1371.

Duncan, Greg J. and Stephen W. Raudenbush. 1999. "Assessing the Effects of Context in Studies of Child and Youth Development". *Educational Psychologist*. 34(1): 29-41.

_____. "Neighborhoods an Adolescent Development: How Can We Determine the Links". Unpublished manuscript. Northwestern University.

Durlauf, Steven N. 2000. "The Memberships Theory of Poverty: The Role of Group Affiliations in Determining Socioeconomic Outcomes". Unpublished manuscript. Department of Economics, University of Wisconsin.

_____. 2001. "A Framework for the Study of Individual Behavior and Social Interactions". Unpublished manuscript. Department of Economics, University of Wisconsin.

_____. 2002 "Groups, Social Influences and Inequality: A Memberships Theory Perspective on Poverty Traps". Unpublished manuscript. Department of Economics, University of Wisconsin.

Ellen IG, Mijanovich T, Dillman K. 2001. "Neighborhood effects on health: exploring the links and assessing the evidence". *Journal of Urban Affairs.* 23(3): 391-408.

Evans WN, Oates W, Schwab RM. 1992. "Measuring peer group effects: a study of teenage behavior". *Journal of Political Economy.* 100:966-91.

Fitzgerald J, Gottschalk P, Moffitt R. 1998. "An analysis of sample attrition in panel data: The Michigan Panel Study of Income Dynamics". *Journal of Human Resources.* 33(2):251-99.

_____. 1998b. "The impact of attrition in the Panel Study of Income Dynamics on intergenerational analysis". *Journal of Human Resources.* 33(2):300-44.

Ganz, Michael Lee. 2001. "Health Threats and Health Practices". Unpublished manuscript. Department of Maternal and Child Health, Harvard School of Public Health.

Geronimus, Arline T. 1996. "Black/White Differences in the Relationship of Maternal Age to Birthweight: A Population-Based Test of the Weathering Hypothesis". *Social Science Medicine*. 42(4): 589-597.

Geronimus, Arline T., John Bound, Timothy A. Waidmann, Cynthia G. Colen and Dianne Steffick. 2001. "Inequality in Life Expectancy, Functional Status, and Active Life Expectancy Across Selected Black and White Populations in the United States". *Demography*. May, 38(2): 227-251.

Griliches, Z. 1979. "Sibling models and data in economics: beginnings of a survey". *Journal of Political Economy.* 87:S37-64.

Hayward, Mark D. and Bridget K. Gorman. "The Long Arm of Childhood: The Influence of Early Life Social Conditions on Men's Mortality". Unpublished manuscript. Population Research Institute, Penn State University.

Kremer, Michael. 1997. "How Much Does Sorting Increase Inequality?" *The Quarterly Journal of Economics*. Feb., 112(1): 115-139.

Kunz, Jim, Marianne E. Page and Gary Solon. 2001. "Are Single-Year Measures of Neighborhood Characteristics Useful Proxies for Children's Long-Run Neighborhood Environment?". Unpublished manuscript. University of Maryland and Joint Center for Poverty Research.

Manski, Charles F. 1993. "Identification of Endogenous Social Effects: The Reflection Problem". *The Review of Economic Studies*. 60(3): 531-542.

Morenoff, Jeff, and John Lynch. 2002. What makes a place healthy? Neighborhood influences on racial/ethnic disparities in health over the life course. Paper prepared for a Workshop on Ethnic Disparities in Aging Health, Washington, DC.

Oreopoulos, Phillip. "The Long-Run Consequences of Living in a Poor Neighborhood". Unpublished manuscript.

Page, Marianne E. and Gary Solon. 1999. "Correlations between Brothers and Neighboring Boys in their Adult Earnings: The Importance of Being Urban". Unpublished manuscript. University of California at Davis.

_____. 2001. "Correlations between Sisters and Neighboring Girls in Their Subsequent Income as Adults". Unpublished manuscript. University of California at Davis.

Rabe-Hesketh, Sophia. 2001. "GLLAMM Manual Technical Report 2001/01". Unpublished manuscript. Department of Biostatistics and Computing, Institute of Psychiatry, University of London.

Raaum, Oddbjørn. 2002. "The Neighbourhood is not what it used to be". Unpublished manuscript. Norwegian School of Economics.

Roberts EM. 1999. "Socioeconomic position and health: The independent contribution of community socioeconomic context". *Annual Review of Sociology*. 25:489-516.

Skinner, Jonathan, D. Staiger, A. Chandra, J. Lee, and M. McClellan. 2002. "Racial Differences in Hospital Quality for the Treatment of Acute Myocardial Infarction: Evidence from the Medicare Population". Unpublished manuscript. Center for Evaluative Clinical Sciences, Dartmouth Medical School.

Solon, Gary, Marianne Page, and Greg Duncan. 2000. "Correlations between neighboring children in their subsequent educational Attainment". *The Review of Economics and Statistics*. 82(3): 383-392.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures". *Journal of Political Economy*. 64: 416-424.

Timberlake, Jeffrey M. 2002. "Racial and Ethnic Inequality in Childhood Exposure to Neighborhood Poverty and Affluence". Unpublished manuscript. Department of Sociology, University of Chicago.

Yen IH, Syme SL. 1999. "The social environment and health: A discussion of the epidoemiologic literature". *Annual Review of Public Health*. 20:287-308.

Table 1
Sibling and Neighbor Sample Sizes for Various Health-related Outcomes

Group	Low Birth Weight	Self-assessed Health in Childhood		Self-assessed Health in Adulthood		Obesity		Cigarette Smoking in Adulthood		High Risk Tolerance
		Males	Females	Males	Females	Males	Females	Males	Females	Males
Sibling Sample:										
# Individuals	3678	565	682	1505	1709	1015	1308	1018	1331	654
# Families	1202	249	286	593	649	428	512	429	520	287
Neighbor Sample:										
#Individuals	2258	544	646	1107	1274	853	1030	459	594	610
# Families	881	373	408	677	725	562	619	307	364	435
# Neighborhoods	327	153	173	252	268	206	235	127	153	173

Table 1a
Neighbor Sample Means for Various Health-related Outcomes

	Low Birth Weight	Health in Childhood		Health in Adulthood		Obesity		Cigarette Smoking in Adulthood		High Risk Tolerance
		Males	Females	Males	Females	Males	Females	Males	Females	Males
Mean Age	--	--	--	38.7	39.3	39.5	38.9	40.0	39.3	36.4
Proportion with health cond'n/behavior	.064	.130	.193	.069	.086	.225	.198	.273	.224	.242

Table 2
Sibling and Neighbor Correlations in Health-related Outcomes
(based on Decomposition of Variance into fraction that lies between neighborhoods, families, and individuals)

	Low Birth Weight	Self-assessed Health in Childhood		Self-assessed Health in Adulthood		Obesity		Cigarette Smoking in Adulthood		High Risk Tolerance
Group		Males	Females	Males	Females	Males	Females	Males	Females	Males
Sibling	0.630	0.286	0.265	0.452	0.517	0.483	0.513	0.560	0.452	0.384
Neighbor	0.082	0.122	0.111	0.257	0.115	0.056	0.096	0.277	0.036	0.183
Adjusted Neighbor	0.075	0.121	0.108	0.242	0.088			0.268	0.013	0.168

Table 2a
Standard Deviation of Neighborhood (Random) Effects for Health-related Outcomes

	Low Birth Weight	Self-assessed Health in Childhood		Self-assessed Health in Adulthood		Obesity		Cigarette Smoking in Adulthood		High Risk Tolerance
		Males	Females	Males	Females	Males	Females	Males	Females	Males
Neighborhood	0.851*	0.790	0.755	1.399***	0.928*	0.811*	1.045*	1.705***	0.522	1.143**

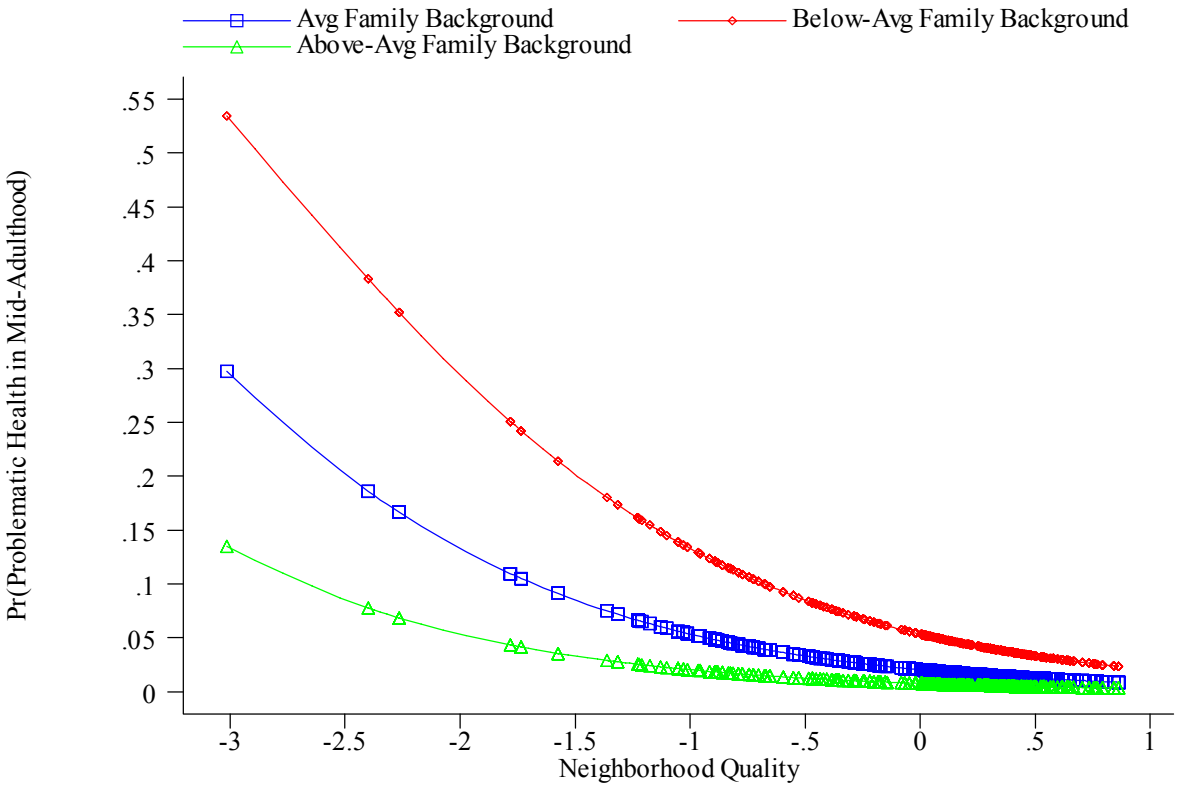
Notes: * p<.10, ** p<.05, and *** p<.01. Based on likelihood ratio tests of random effects.

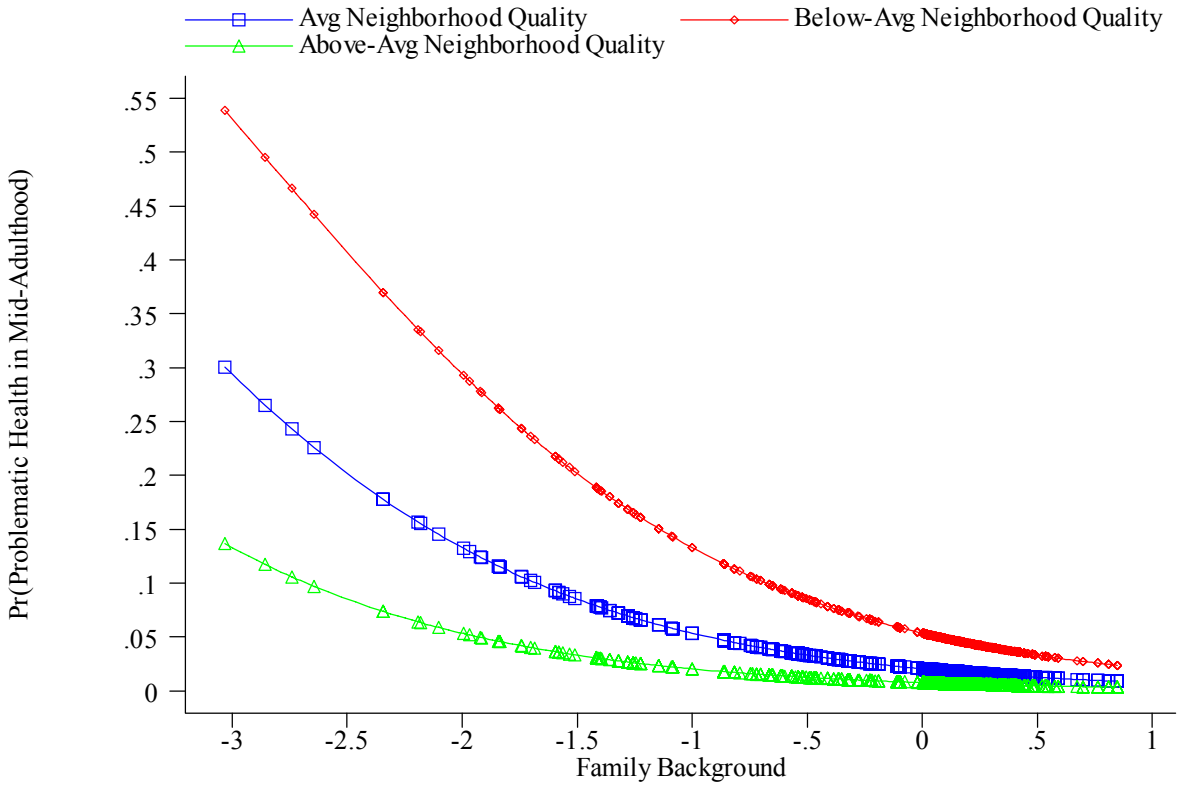
Table 3
Multilevel Discrete-Time Hazard of Onset of Adolescent Cigarette Use, separately by gender

	Boys Coef. Est.	Girls Coef. Est.
Age		
10	---	1.3945
		(1.2389)
11	---	0.6607
		(1.4270)
12	2.9085**	3.0683***
	(1.1999)	(.9575)
13	3.6212***	3.2075***
	(.9335)	(1.0564)
14	4.4285***	3.4914***
	(1.0643)	(1.0967)
15	3.8398***	3.9296***
	(1.2030)	(1.0591)
16	5.0119***	4.3875***
	(1.0997)	(1.0662)
17	4.0889***	4.1199***
	(1.2504)	(1.1368)
18	5.9910***	4.9547***
	(1.2115)	(1.0937)
Year	-0.0506	0.0099
	(.0339)	(.0379)
Constant	-4.3399	8.0504***
	(1.9334)	(2.6661)
Unmeasured (Std Dev)		
Neighborhood Component	.9154***	0.6104
Family Component	1.0046***	1.2292***
Log-Likelihood	-306.0723	-404.2606
Sample Size:		
# Observations	2523	2915
# Individuals	272	317
# Families	191	215
# Neighborhoods	77	91
# Failures	84	99

Notes: * p<.10, ** p<.05, and *** p<.01.
Robust standard errors are in parentheses.

Figures 1 and 2: Prob of Problematic Health in Mid Adulthood for Males over Range of Family and Neighborhood Effects





Figures 3 and 4:
Prob of Low Birth Weight over Range of Family and Neighborhood Effects